DNA Triple Helix Formation at Target Sites Containing Several Pyrimidine Interruptions: Stabilization by Protonated Cytosine or 5-(1-Propargylamino)dU†

Darren M. Gowers,‡,§ Jeevan Bijapur,|| Tom Brown,|| and Keith R. Fox*,‡

Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K., and Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Received May 20, 1999; Revised Manuscript Received August 4, 1999

ABSTRACT: DNase I footprinting has been used to study the formation of parallel triplexes at oligopurine target sequences which are interrupted by pyrimidines at regular intervals. TA interruptions are targeted with third strand oligonucleotides containing guanine, generating G·TA triplets, while CG base pairs are targeted with thymine, forming T·CG triplets. We have attempted to optimize the stability of these complexes by varying the base composition and sequence arrangement of the target sites, and by replacing the third strand thymines with the positively charged analogue 5-(1-propargylamino)dU (U^P). For the target sequence $(AAAT)_5AA$, in which pyrimidines are positioned at every fourth residue, triplex formation with TG-containing oligonucleotides is only detected in the presence of a triplex-binding ligand, though stable triplexes were detected at the target site $(AAAAAT)_3AAAA$. Triplex stability at targets containing pyrimidines at every fourth residue is increased by introducing guanines into the duplex repeat unit using the targets $(AGAT)_5AA$ and $(ATGA)_5AA$. In contrast, placing C^+·GC triplets on the 5′-side of G·TA, using the target $(AGTA)_5TT$, produces complexes of lower stability. We have attempted further to increase the stability of these complexes by using the positively charged thymine base analogue U^P, and have shown that $(TU^PTG)_5TT$ forms a more stable complex with target $(AAAT)_5AA$ than the unmodified third strand, generating a footprint in the absence of a triplex-binding ligand. Triplex formation at $(AGTA)_5AA$ is improved by using the modified oligonucleotide $(TCGU^P)_5TT$, generating a complex in which the charged triplets C^+·GC and U^P·AT alternate with uncharged triplets. In contrast, placing U^P·AT triplets adjacent to C^+·GC, using the third strand oligonucleotide $(U^PCGT)_5TT$, reduces triplex formation, while the third strand with both substitutions, $(U^PCGU^P)_5TT$, produces a complex with intermediate stability. It appears that, although adjacent U^P·AT triplets form stable triplexes, placing U^P·AT adjacent to C^+·GC is unfavorable. Similar results were obtained with fragments containing CG inversions within the oligopurine tract, though triplexes at $(AAAAAC)_3AA$ were only detected in the presence of a triplex-binding ligand. Placing C^+·GC on the 5′-side of T·CG triplets also reduces triplex formation, while a 3′-C^+·GC produces complexes with increased stability.

Intermolecular DNA triple helices are formed by the sequence-specific binding of an oligonucleotide in the major groove of duplex DNA where it makes contacts with substituents on the purine bases (*1*, *2*). Two types of triple helix have been characterized which differ according to the orientation of the third strand. Pyrimidine-rich oligonucleotides bind in a parallel orientation with respect to the duplex purine strand, forming T·AT and C^+·GC triplets (*3*−*5*), while purine-rich third strands bind in an antiparallel orientation, generating G·GC, A·AT, or reverse Hoogsteen T·AT triplets (*6*−*9*). Conditions of low pH are generally required for generating parallel triplexes in order to achieve protonation of the third strand cytosines. Since triplex-forming oligonucleotides bind to DNA with considerable sequence selectivity, they have been proposed as antigene agents for treating viruses or cancer, or as tools in molecular biology (*10*−*13*).

A major problem in the use of triplexes for achieving sequence-specific recognition of DNA is that these complexes are usually restricted to homopurine tracts, since there is no simple means of recognizing pyrimidine bases (*14*, *15*). Several base analogues have been tested as candidates for overcoming this limitation (*14*, *15*), though without much success to date. However, two triplets using natural bases have been suggested in the parallel motif, namely, G·TA (*16*−*27*) and T·CG (*19*, *25*, *27*, *28*). The structures of these triplets are shown in Figure 1. Although these triplets are specific, their stability is much lower than T·AT or C^+·GC since the third strand base only makes one hydrogen bond contact with the duplex pyrimidine. In addition, their formation is accompanied by local distortions in backbone

† This work was supported by grants from the Cancer Research Campaign.

* Corresponding author: Dr. Keith R. Fox, Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K. Telephone: +1703-594374. Fax: +1703-594459. Email: KRF1@soton.ac.uk.

‡ Division of Biochemistry and Molecular Biology, University of Southampton.

§ Present address: Department of Biochemistry, University of Bristol, School of Medical Sciences, University Walk, Bristol BS8 1TD, U.K.

|| Department of Chemistry, University of Southampton.

10.1021/bi9911637 CCC: $18.00
Published on Web 09/24/1999

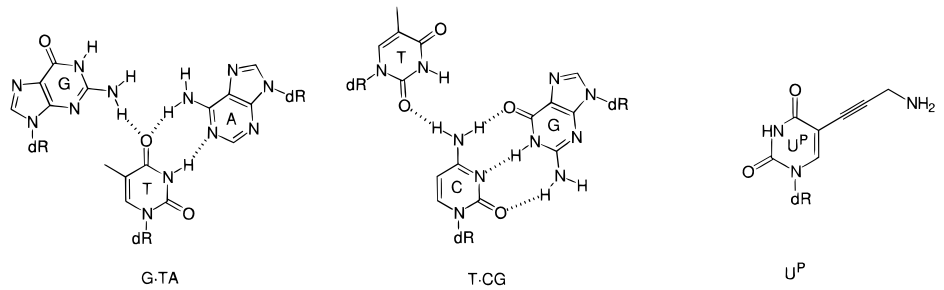

FIGURE 1: Chemical structures of the G·TA triplet, the T·CG triplet, and the thymine base analogue 5-(1-propargylamino)dU (U^P).

geometry, since these triplets are not isomorphous with T·AT and C^+·GC. As a result, additional stabilizing factors such as triplex-binding ligands or tethered intercalating agents are required to improve the affinity of triplexes which contain these triplets (*24*−*26*, *29*−*31*). Increasing the stability of oligonucleotides containing these triplets, without the need for these agents, is therefore of considerable interest.

At the center of a parallel triplex, each additional G·TA causes a 30-fold decrease in third strand affinity (*25*), though complexes containing up to three adjacent G·TA triplets can be formed in the presence of a triplex-binding ligand (*25*). The G·TA triplet is more stable when flanked by T·AT than C^+·GC (*17*), an effect which is due to the presence of an additional hydrogen bond to thymine in the adjacent T·AT triplet (*21*−*23*, *33*). We have previously examined the possibility of targeting duplex regions of $(AT)_n$ with GT-containing oligonucleotides, generating blocks of alternating G·TA and T·AT triplets (*24*, *26*). Although these structures are not stable, even in the presence a triplex-binding ligand, triplex formation can be induced by attaching this region to an adjacent block of consecutive T·AT triplets (*24*, *26*). Similar complexes containing a few C^+·GC triplets in the anchoring tail are more stable and form in the presence of Mg^{2+}, without addition of a stabilizing ligand (*26*). This is consistent with the observation that C^+·GC imparts a greater stability to triplex structures than T·AT (*34*−*37*), and raises the possibility that one means of facilitating triplex formation at pyrimidine inversions is to increase the stability of the surrounding canonical T·AT and C^+·GC triplets.

The stable interaction of thymine with a CG base pair in a parallel triplex was first proposed by Yoon et al. (*27*). Since this triplet contains a single hydrogen bond between O2 of thymine and the exocyclic N4 of cytosine, it is weaker than the canonical T·AT triplet. In most reports, T·CG appears to be weaker than G·TA (*32*, *33*). As with G·TA, up to three consecutive T·CG triplets can be accommodated within a DNA triplex if the interaction is stabilized by a triplex-binding ligand (*25*).

One approach for overcoming the low affinity of G·TA- or T·CG-containing triplexes, with which this paper is concerned, is to compensate for the local loss of binding energy at TA or CG by increasing the stability of surrounding canonical triplets. One means of achieving this is to increase the number of isolated C^+·GC triplets, since this triplet has previously been shown to be more stable than T·AT (*34*−*37*), though adjacent C^+·GC triplets are destabilizing. An alternative approach is to increase the stability of the T·AT triplet, using charged thymine analogues. 5-(1-Propargylamino)dU (U^P) (Figure 1) is a novel base analogue of uracil which is protonated near physiological pH and forms more stable triplets at AT base pairs than thymine (*38*). This improved stability is thought to arise from stacking of the propynyl residues within the major groove, as well as interaction of the terminal amino group with the phosphodiester backbone. The work described in this paper uses C^+·GC and U^P·AT triplets to stabilize triplexes containing either G·TA or T·CG triplets.

In previous work we have shown that $(AT)_n$ tracts cannot be recognized by third strands of the type $(TG)_n$ even in the presence of a triplex-binding ligand (*24*, *26*). We reasoned that increasing the number of canonical triplets between each G·TA (or T·CG) triplet, and including some C^+·GC triplets, would increase oligonucleotide affinity. We have therefore prepared DNA fragments containing sequences of the type $(R_3Y)_n$ and used these as targets for triple helix formation. These sequences were designed to examine how different arrangements of triplets affect the stability of triplexes targeted at sequences containing up to 25% pyrimidine interruptions. We have attempted to optimize the stability of these complexes by varying the base composition and sequence arrangement of the target sites, and by replacing the third strand thymines with the positively charged analogue 5-(1-propargylamino)dU (U^P). These 22mer sites, which are summarized in Figure 2 along with the third strand oligonucleotides examined, contained the sequences $(A_3Y)_5$-AA, $(AGAY)_5AA$, $(AGYA)_5AA$, and $(AYGA)_5AA$ (Y = T or C). These were targeted with third strands containing natural bases or U^P residues substituted in place of T. In some instances, the complexes were further stabilized by addition of a 10 μM sample of a naphthylquinoline triplex-binding ligand (*39*−*42*).

MATERIALS AND METHODS

Chemicals and Enzymes. Oligodeoxynucleotides were purchased from Oswel DNA Service. Sequences containing U^P residues were prepared as previously described (*38*). *Bam*HI-cut alkaline phosphatase-treated pUC18, DNA ligase, and radiochemicals were from Amersham Pharmacia Biotech Ltd. Bovine DNase I was purchased from Sigma and stored frozen at 7200 units/mL. Restriction enzymes and reverse transcriptase were purchased from Promega. The naphthylquinoline triplex-binding ligand (*39*−*42*) was stored at a concentration of 20 mM in dimethyl sulfoxide at −20 °C and was a gift from Dr. Lucjan Strekowski, Department Chemistry, Georgia State University.

DNA Sequences. Complementary oligonucleotides were treated with polynucleotide kinase and annealed prior to ligation into *Bam*HI-cut pUC18. Following transformation of calcium-permeabilized *E. coli* TG2, successful clones were picked from agar plates containing ampicillin IPTG and

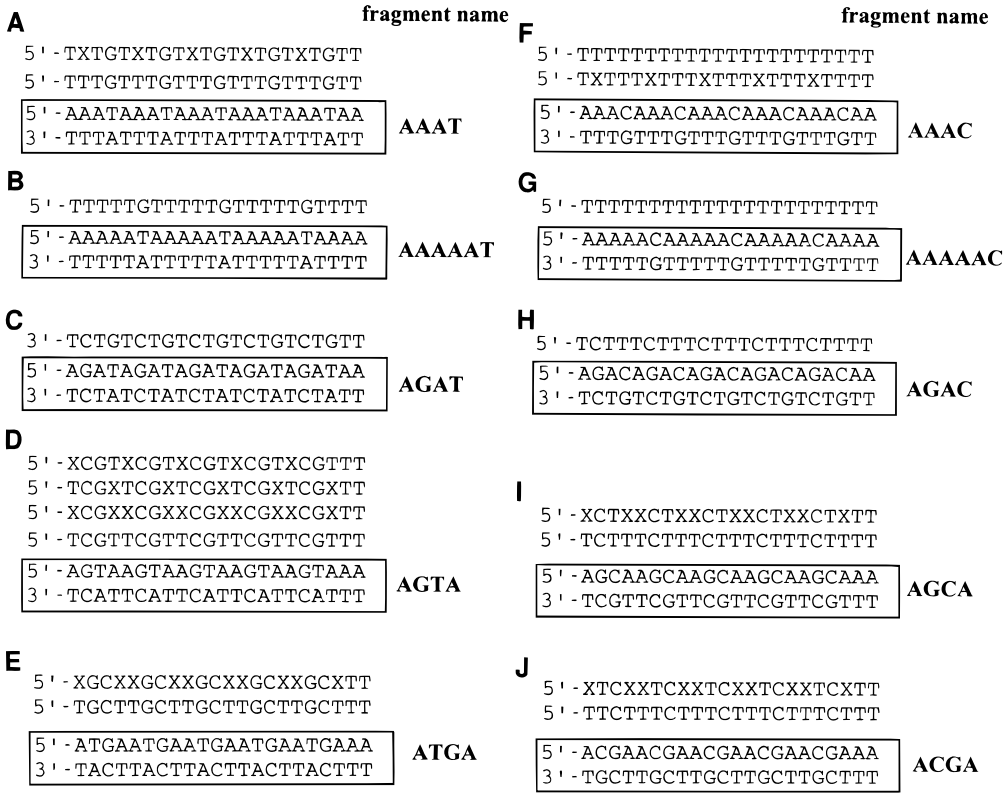

FIGURE 2: Sequences of the 10 triplex target sites used in this work (boxed) together with their respective third strand oligonucleotides. X denotes the modified base U^p. The short fragment name is shown alongside each target sequence. (A) Target $(AAAT)_4AA$, (B) target $(AAAAAT)_3AAAA$, (C) target $(AGAT)_4AA$, (D) target $(AGTA)_4AA$, (E) target $(ATGA)_4AA$, (F) target $(AAAC)_4AA$, (G) target $(AAAAAC)_3AAAA$, (H) target $(AGAC)_4AA$, (I) $(AGCA)_4AA$, (J) $(ACGA)_4AA$. All fragments were cloned into the *Bam*HI site of pUC18. Fragments AAAT, AGAT, AGTA, ATGA, AAAAAC, AGCA, and ACGA were oriented so that labeling at the 3′-end of the *Hin*dIII site visualized the purine-rich strand while for fragments AAAAAT, AAAC, and AGAC the pyrimidine-rich strand was visualized.

X-Gal as white colonies. Clones were sequenced using a T7 dideoxy sequencing kit (Amersham Pharmacia Biotech Ltd.). The sequences of the 10 different target sites are presented in Figure 2, along with their respective third strand oligonucleotides. The inserts were oriented so that labeling the 3′-end of the *Hin*dIII site visualized the purine-rich strand of all sequences except AAAAAT, AAAC, and AGAC, for which the pyrimidine-rich strand was visualized.

DNA Fragments. Plasmids containing the cloned inserts were digested with *Hin*dIII and *Sac*I before labeling at the 3′-end of the *Hin*dIII site with [α-^{32}P]dATP using reverse transcriptase. The fragments were separated from the remainder of the plasmid on 6% (w/v) nondenaturing polyacrylamide gels. The labeled DNA was eluted from the gel and dissolved in 10 mM Tris-HCl, pH 7.5, containing 0.1 mM EDTA at a concentration of 10−20 cps/μL as measured on a hand-held Geiger counter (approximately 10 nM).

DNase I Footprinting. Radiolabeled DNA (1.5 μL), oligonucleotide (1.5 μL), and 1.5 μL of buffer or triplex-binding ligand (30 μM) were mixed to give final third strand concentrations between 100 and 1×10^{-3} μM. For experiments at pH 5.5, the oligonucleotide and triplex-binding ligand were dissolved in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$; for experiments at pH 7.5, the buffer used was 10 mM Tris-HCl, pH 7.5, containing 10 mM $MgCl_2$ and 50 mM NaCl. The complexes were left to equilibrate for at least 3 h at 20 °C. Digestion was started by adding 2 μL of DNase I, dissolved in 20 mM NaCl, 2 mM $MnCl_2$, and 2 mM $MgCl_2$, and terminated after 1 min by adding 4 μL of 80% formamide, containing 10 mM EDTA and 0.1% (w/v) bromophenol blue.

Electrophoresis. Products of digestion were separated on 10% (w/v) denaturing polyacrylamide gels (National Diagnostics) containing 8 M urea. Electrophoresis conditions were 1500 V for about 2 h. Gels were fixed in 10% (v/v) acetic acid before being dried at 80 °C for 1 h and exposed overnight to autoradiography film at −70 °C using an intensifying screen. Bands in each digestion pattern were assigned by comparison with Maxam−Gilbert markers specific for purines.

RESULTS

Complexes Containing G·TA Triplets

We have previously shown that blocks of alternating G·TA and T·AT triplets are not stable, even in the presence of a triplex-binding ligand (*24*, *26*), and that $(GT)_n$ third strands cannot form stable triplexes at $(AT)_n$ target sites unless they are tethered to another stable triplex such as $(T{\cdot}AT)_n$. We have therefore investigated how the proportion of G·TA triplets affects triplex stability by increasing the number of intervening T·AT triplets, using the 22mer target sites $(A_3T)_5$-AA and $(A_5T)_3AAAA$.

Triplex Formation at $(AAAT)_5AA$ and $(AAAAAT)_3AAAA$. Fragment AAAT contains a target site which is designed to generate a triplex with the third strand oligonucleotide $(TTTG)_5TT$ in which five G·TA triplets are each separated by three T·AT triplets (see Figure 2A). The first panel of

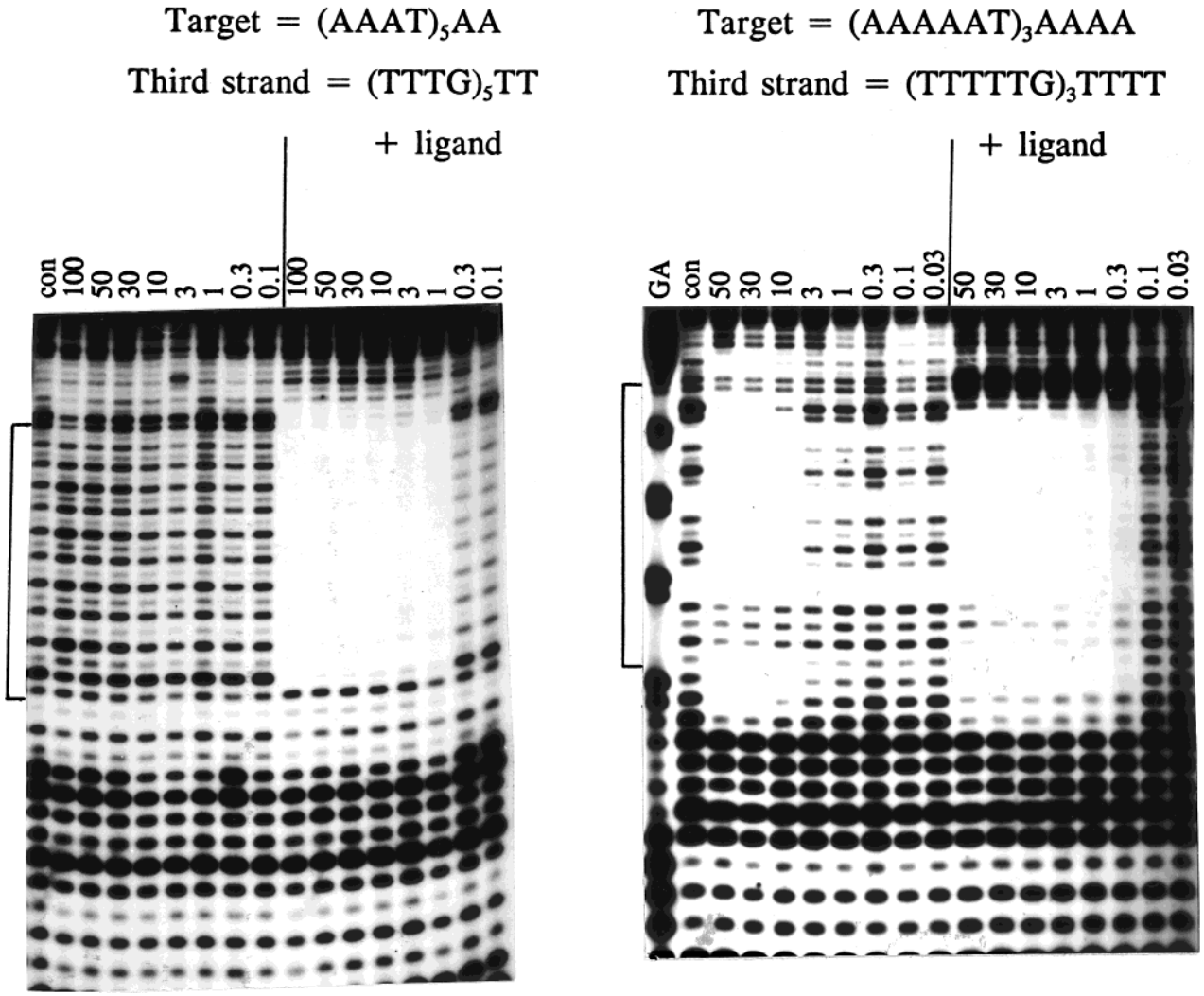

FIGURE 3: DNase I footprinting patterns showing the interaction of oligonucleotides $(TTTG)_5TT$ and $(TTTTTG)_3TTTT$ with fragments AAAT and AAAAAT, respectively. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. In each panel, the right-hand lanes contained 10 μM naphthylquinoline triplex-binding ligand. Reactions were performed in 10 mM Tris-HCl, pH 7.5, containing 50 mM NaCl and 10 mM $MgCl_2$. Tracks labeled con are controls performed in the absence of added oligonucleotide or ligand. The track labeled GA is a Maxam−Gilbert marker specific for purines. The square brackets show the position of the target sites.

Figure 3 shows the results of DNase I footprinting experiments in the presence and absence of both the third strand oligonucleotide and a naphthylquinoline triplex-binding ligand at pH 7.5 in the presence of 10 mM $MgCl_2$. No interaction with $(TTTG)_5TT$ can be seen in the absence of ligand (left-hand lanes), even at the highest oligonucleotide concentration (100 μM). In the presence of 10 μM ligand (right-hand lanes), a clear footprint is evident covering the entire target site, which persists to an oligonucleotide concentration of about 1 μM. This large increase in binding affinity is in contrast to the inability of the ligand to stabilize the formation of similar triplexes at alternating AT tracts, and is presumably assisted by its interaction with adjacent T·AT triplets.

The second panel of Figure 3 shows the results of similar experiments with fragment AAAAAT, designed to form a triplex with $(T_5G)_3T_4$ in which the three G·TA triplets are each separated by five T·AT triplets (Figure 2B). In the absence of the ligand (left-hand lanes), a footprint can be seen at the target site at oligonucleotide concentrations of 10 μM and above. Some cleavage products are still evident toward the 3′-end (lower) of the target site, suggesting that the third strand may be fraying at this end. On addition of 10 μM naphthylquinoline triplex-binding ligand, the interaction is enhanced, and the footprint persists to an oligonucleotide concentration of about 0.3 μM. The footprint in the presence of the ligand is shorter at the 5′-end (upper) of the target site; this effect has been previously noted on the pyrimidine (not the purine) strand of ligand-induced triplexes (*41*), and appears to represent an overestimation of the footprinting site size with the oligonucleotide alone. When the DNA fragment was labeled on the purine-rich strand (not shown), the ligand did not affect the size of the footprint, but showed a similar increase in stability in the presence of the ligand. The increased stability of triplexes formed at AAAAAT compared with AAAT is probably a combination of the reduction in the number of G·TA triplets as well as an increase in the length of the separating A_n·T_n tracts.

Triplex Formation at (AGAT)5AA, (AGTA)5AA, and (ATGA)5AA. Several studies have shown that isolated C^+·GC triplets impart a greater triplex stability than T·AT (*34*−*37*). We therefore attempted to improve the stability of triplexes at sites similar to AAAT, in which every fourth residue on the purine-rich strand is T, by replacing some of

the target AT pairs with GC. Since it has previously been suggested that the G·TA triplet is more stable when flanked by T·AT than C^+·GC (*19*), we used three different sequences in which the G·TAs are flanked on both sides by T·AT (fragment AGAT, Figure 2C), and with C^+·GC on either the 5′-side (AGTA, Figure 2D) or the 3′-side (ATGA, Figure 2E). These experiments were all performed at pH 5.5, necessary to ensure protonation of the third strand cytosines.

Figure 4 (left-hand panel) shows the interaction of oligonucleotide $(TCTG)_5TT$ with fragment AGAT in the absence or presence of the triplex-binding ligand. In this instance, a clear footprint is evident in the absence of the ligand (left-hand lanes), which persists to an oligonucleotide concentration of 1 μM. The introduction of these five isolated C^+·GC triplets has clearly increased triplex affinity compared to fragment AAAT, for which no triplex formation was observed in the absence of the ligand. Addition of the naphthylquinoline triplex-binding ligand (right-hand lanes) has little effect on the footprinting pattern. The inability of the ligand to potentiate triplex formation at this site is consistent with the suggestion that it binds best to adjacent T·AT triplets (*42, 43*), since this fragment contains only one ApA step at the 3′-end of the target site.

Figure 4 (center panel) shows the results of similar experiments with fragment AGTA, examining triplex formation with $(TCGT)_5TT$. In the absence of ligand (left-hand lanes), there is some protection from DNase I cleavage, though weak bands are still evident within the target site, even at the highest oligonucleotide concentration. The lack of a complete footprint suggests that this triplex is weaker than that formed with AGAT and confirms that G·TA is less stable when flanked by C^+·GC on the 5′-side. On addition of the triplex-binding ligand (right-hand lanes), the intensity of these weak bands is reduced, and the footprint continues to 0.3 μM. The apparent increase in affinity on addition of ligand is again consistent with the suggestion that it interacts with adjacent T·AT triplets, since this site contains several ApA steps.

The third panel of Figure 4 shows the results of similar experiments with fragment ATGA, examining the interaction with $(TGCT)_5TT$, generating a complex in which the G·TA triplets are flanked by C^+·GC on the 3′-side. In the absence of ligand, a clear footprint is evident which persists to an oligonucleotide concentration of 0.3 μM, which is comparable to that seen with fragment AGAT. This suggests that the stability of G·TA is not affected by the presence of a C^+·GC triplet on the 3′-side. In the presence of 10 μM triplex-binding ligand (right-hand lanes), the footprint also continues to about 0.3 μM, and this is accompanied by enhanced cleavage at the upper end (3′-end) of the target site.

Replacement of T with U^P. Since the inclusion of some C^+·GC triplets appears to increase the stability of G·TA-containing triplexes, we reasoned that charged analogues of thymine might also potentiate the formation of stable complexes. We have recently shown that replacement of T with 5-(1-propargylamino)dU (U^P) greatly increases triplex stability in a pH-sensitive fashion (*38*), probably by virtue of a favorable interaction of the charged side group with the phosphodiester backbone. We have therefore studied the binding of some U^P-containing oligonucleotides to the sequences described above.

The first panel of Figure 5 shows the interaction of oligonucleotide $(TU^PTG)_5TT$ with fragment AAAT, replacing the central T·AT triplets of each repeating unit with U^P·AT. It can be seen that this modified oligonucleotide produces a clear footprint which persists to an oligonucleotide concentration of about 1 μM. Although this experiment was performed at pH 5.5, to ensure full protonation of the propargylamino groups, a similar footprint is also produced at pH 7.5 (not shown) which requires only slightly higher oligonucleotide concentrations (3 μM). The position of the U^P·AT triplets in the sequence is similar to the C^+·GCs in the complex with AGAT, and clearly shows that addition of positively charged residues enhances triplex stability compared to the unmodified oligonucleotide which does not produce a DNase I footprint at this target site without the addition of a triplex-binding ligand. Since the charged side group of this base analogue is presumed to interact with the phosphodiester backbone, we investigated whether inclusion of this base might alter the magnesium dependency of triplex formation. However, we find that triplex formation still requires millimolar concentrations of magnesium (not shown), presumably because substitution of T with U^P only introduces 1 additional positive charge for every 4 triplets (12 phosphates).

The central three panels of Figure 5 show DNase I footprints for the interaction of three U^P-containing oligonucleotides with fragment AGTA at pH 5.5. In these modified oligonucleotides, the U^P residues have been positioned so as to recognize either the AT base pair on the 3′-side of each G·TA triplet $[(TCGU^P)_5TT]$ or the AT base pair on the 5′-side of the C^+·GC triplet $[(U^PCGT)_5TT]$, or both AT base pairs $[(U^PCGU^P)_5TT]$. The results with $(TCGU^P)_5$-TT, in which the U^P·AT triplets are sandwiched between G·TA and T·AT, show a clear footprint at the target site which persists to about 0.3 μM oligonucleotide. As described above, the unmodified oligonucleotide $(TCGT)_5TT$ does not fully protect from cleavage at this target. It therefore appears that substitution of the second T by U^P, generating a complex in which charged triplets (U^P·AT and C^+·GC) are present at every other position, increases triplex stability relative to the unmodified oligonucleotide. In contrast, oligonucleotide $(U^P$-$CGT)_5TT$, in which the U^P·AT triplets are placed between T·AT and C^+·GC, produces very little protection from DNase I cleavage, though there is some attenuation of band intensity at the highest oligonucleotide concentrations (50 and 30 μM). This oligonucleotide appears to bind less well than the unmodified oligonucleotide $(TCGT)_5TT$, suggesting that juxtaposition of the two charged triplets is unfavorable. The fourth panel of Figure 5 shows the results for oligonucleotide $(U^PCGU^P)_5TT$, containing both U^P substitutions. This oligonucleotide produces a clear footprint at the target site with concentrations of 10 μM and above; this is better than the unmodified oligonucleotide but inferior to $(TCGU^P)_5TT$. This complex contains blocks of three charged triplets (U^P·AT, U^P·AT, C^+·GC), separated by single G·TA triplets. The intermediate stability of this structure with three consecutive charged triplets is surprising, since adjacent C^+·GC and U^P·AT triplets are poorly tolerated, but is consistent with our previous observation that adjacent U^P·AT triplets are not destabilizing (*38*).

The final panel of Figure 5 shows the interaction of $(U^P$-$GCU^P)_5TT$ with fragment ATGA, revealing a footprint at

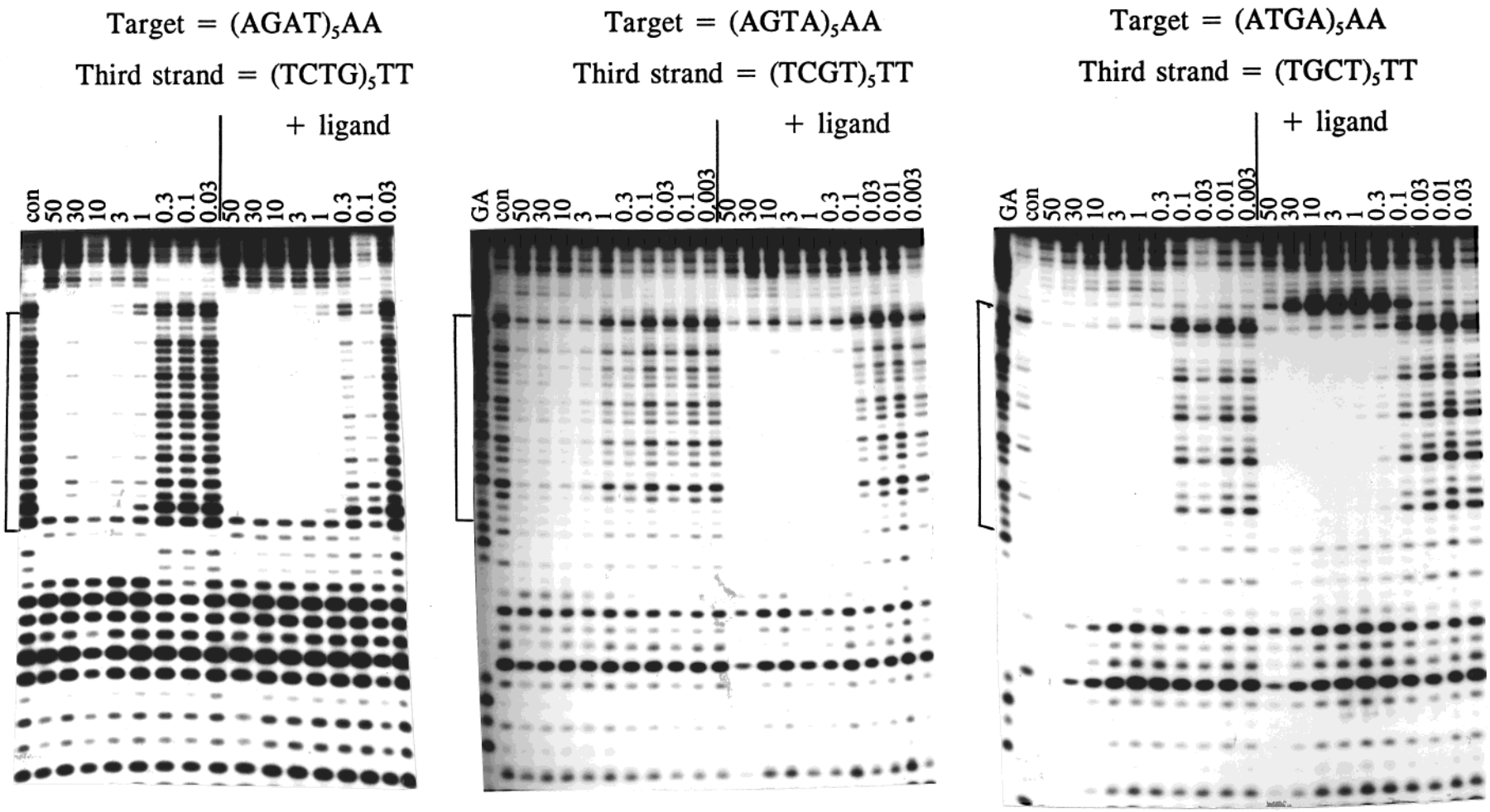

FIGURE 4: DNase I footprinting patterns showing the interaction of oligonucleotide $(TCTG)_5TT$ with fragment AGAT (left-hand panel), $(TCGT)_5TT$ with fragment AGTA (center panel), and $(TGCT)_5TT$ with fragment ATGA (right-hand panel). Oligonucleotide concentrations (μM) are shown at the top of each gel lane. In each panel, the right-hand lanes contained 10 μM naphthylquinoline triplex-binding ligand. Reactions were all performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Tracks labeled con are controls performed in the absence of added oligonucleotide or ligand. The tracks labeled GA are Maxam−Gilbert markers specific for purines. The square brackets show the positions of the target sites.

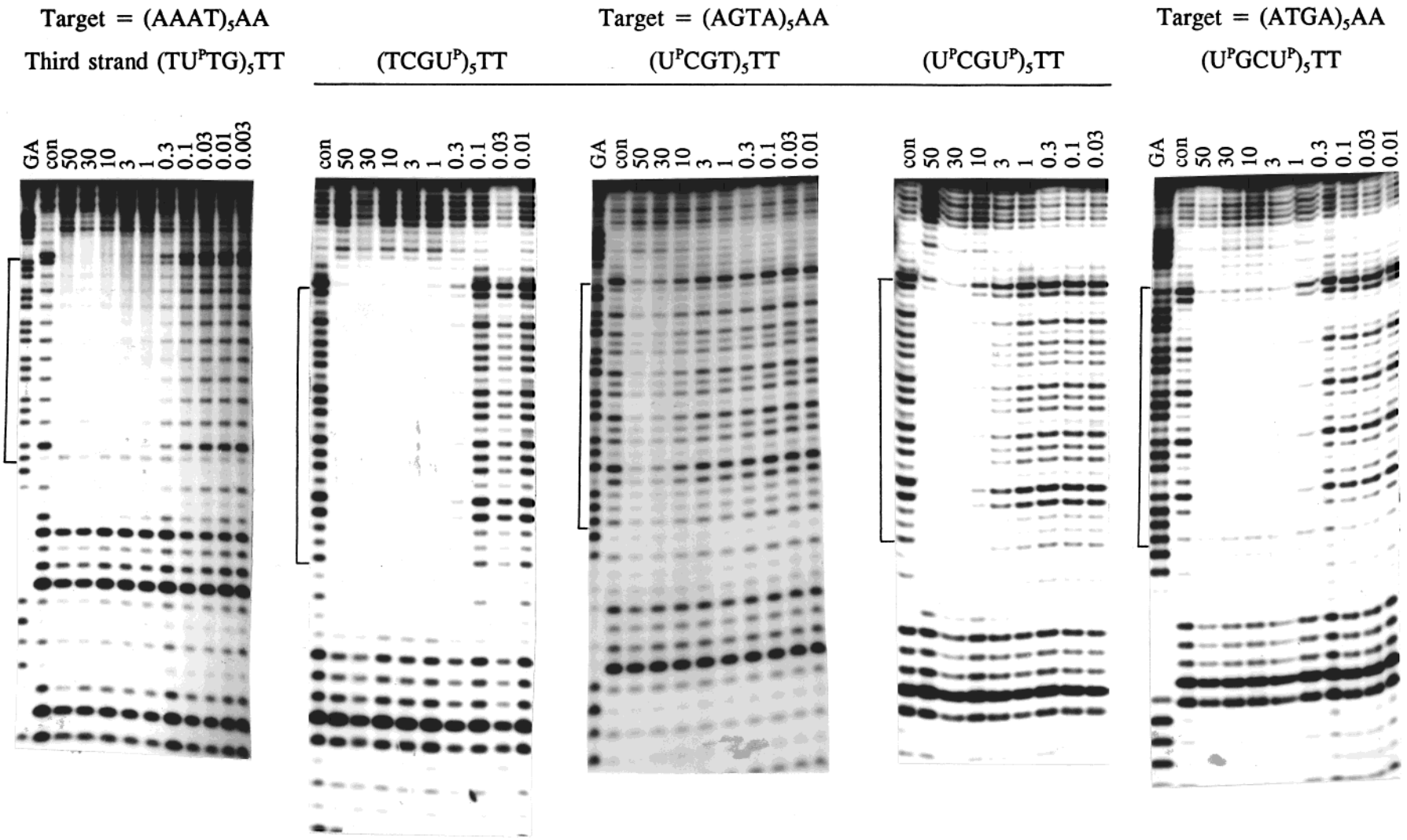

FIGURE 5: DNase I footprinting patterns showing the interaction of U^P-containing oligonucleotides with fragments AAAT (left-hand panel), AGTA (center three panels), and ATGA (right-hand panel). Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Reactions were all performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Tracks labeled con are controls performed in the absence of added oligonucleotide or ligand. The tracks labeled GA are Maxam−Gilbert markers specific for purines. The square brackets show the positions of the target sites.

the target site which persists to about 1 μM oligonucleotide. This doubly substituted oligonucleotide binds no better, and possibly worse, than the unmodified oligonucleotide $(TGCT)_5TT$. This lack of positive effect presumably arises because the beneficial effect of the U^P substitution is canceled out by the close proximity to a C^+·GC triplet. The stability of this complex is similar to the analogous triplex formed between $(U^PCGU^P)_5TT$ and AGTA.

Since replacement of T with U^P stabilizes some complexes which contain G·TA triplets, we attempted to generate complexes containing blocks of alternating G·TA and U^P·AT triplets by targeting duplex regions of $(AT)_n$ with $(U^PG)_n$. Previous studies have shown that $(AT)_{11}$ is not bound by $(TG)_{11}$, even in the presence of a triplex-binding ligand (*26*). In similar experiments to those described above, fragments k2 and k2rev (*26*), which contain $(AT)_{11}$ tracts, were targeted with $(U^PG)_{11}T$ in the presence and absence of 10 μM triplex-binding ligand (not shown). No footprints were detected under any conditions, suggesting that although U^P·AT is more stable than T·AT, this base analogue is not able to facilitate the formation of blocks of alternating G·TA and U^P·AT triplets.

Complexes Containing T·CG Triplets

Figure 6 shows the results of similar experiments with fragments in which the oligopurine tracts are interrupted by cytosines, which we have attempted to recognize by forming T·CG triplets. Although recognition of C by a third strand T is specific, the T·CG triplet is much weaker than the canonical triplets T·AT and C^+·GC and is thought to be weaker than G·TA (*32*−*33*).

Triplex Formation at AAAC. The first panel of Figure 6 shows the interaction of oligonucleotide $(TTTT)_5TT$ with fragment AAAC, designed to form a complex containing five T·CG triplets which are each separated by three T·AT triplets (Figure 2F). As seen with the analogous G·TA-containing triplex (Figure 3), no footprint is visible with the oligonucleotide alone. On addition of 10 μM of the triplex-binding ligand, a clear footprint is evident which persists to an oligonucleotide concentration of about 1 μM. This is accompanied by a band of enhanced cleavage at the upper (5′) end of the target site and is similar to that frequently observed on the pyrimidine-rich target sequence in the presence of triplex-binding ligands (*41*). Since third strand Ts can recognize either AT or CG base pairs, there is considerable redundancy in the binding of this oligonucleotide which might form several related slipped structures. However, it should be noted that the footprint covers the predicted target sequence and does not overlap into surrounding regions.

We attempted to enhance the binding of this oligonucleotide by introducing U^P residues into the third strand, using the oligonucleotide $(TU^PTT)_5TT$, forming a complex which is analogous to that formed between fragments AAAT and $(TU^P\ TG)_5TT$. However, no footprint was apparent with this modified oligonucleotide alone, though slight band attenuation was seen inside the target region at concentrations of 50 and 30 μM (not shown). This result is in contrast to that seen with the equivalent G·TA-containing complex, which forms a clear footprint with the modified oligonucleotide $(TU^PTG)_5TT$ (Figure 5, left-hand panel).

We also increased the proportion of T·AT triplets by targeting the fragment AAAAAC with the third strand oligonucleotide T_{22}, attempting to form a triplex in which the T·CG triplets are each separated by five T·AT triplets (Figure 2G). In contrast to the results with the analogous sequence AAAAAT, no footprint was observed in the absence of added ligand, even at 50 μM oligonucleotide (not shown). On addition of 10 μM naphthylquinoline triplex-binding ligand, the interaction is enhanced, and a footprint is induced which persists to 0.1 μM oligonucleotide.

Triplex Formation at AGAC, AGCA, and ACGA. As described above for the G·TA-containing triplexes, we attempted to increase the stability of these T·CG-containing triplexes by introducing C^+·GC triplets into the structure. Since the stability of G·TA is known to be affected by the flanking triplets, we used three different sequences in which the T·CGs are flanked on both sides by T·AT (fragment AGAC, Figure 2H), with C^+·GC on the 5′-side (AGCA, Figure 2I) and with C^+·GC on the 3′-side (ACGA, Figure 2J). These experiments were all performed at pH 5.5, necessary to ensure protonation of the third strand cytosines.

The second panel of Figure 6 shows the interaction between oligonucleotide $(TCTT)_5TT$ and fragment AGAC. This is designed to generate a complex in which the T·CG triplets should be flanked by T·AT on both sides. In the absence of the triplex-binding ligand, a clear footprint is evident at the highest oligonucleotide concentrations (30 and 50 μM), though bands within the target site are still attenuated at concentrations as low as 1 μM. It is clear that this triplex has a higher affinity than that formed at AAAC and confirms that the isolated C^+·GC triplets have increased the stability. Within the concentration range 1−10 μM, bands are less well protected toward the bottom (3′-end) of the footprint, suggesting that the oligonucleotide may be fraying from this end of the target site. In the presence of 10 μM naphthylquinoline triplex-binding ligand, the footprint is much clearer, but still persists to a concentration of about 1 μM. This footprint is accompanied by enhanced cleavage at the upper (5′) end of the target site, as previously noted for the pyrimidine-rich strand in the presence of triplex-binding ligands. The lack of effect of the ligand on the lowest oligonucleotide concentration which affects the digestion pattern is presumably because it binds best to blocks of T·AT triplets and this fragment, like AGAT, contains only one ApA step, at the 3′-end of the target site.

The third panel of Figure 6 shows the interaction between oligonucleotide $(TCTT)_5TT$ and fragment AGCA, designed to form a triplex in which the T·CG triplets are flanked by C^+·GC on the 5′-side and T·AT on the 3′-side. (The third strand oligonucleotide used with this target is the same as that used with fragment AGAC since both CG and AT base pairs are recognized by T.) In this instance, no footprint is observed in the absence of added ligand, suggesting that placing a C^+·GC on the 5′-side of T·CG is not favorable. A clear footprint is evident in the presence of 10 μM of the triplex-binding ligand, which persists to an oligonucleotide concentration of about 0.3 μM. Stabilization by the ligand is not surprising since this ligand contains several ApA steps. Substitution of U^P·AT for T·AT triplets using the oligonucleotide $(U^PCTU^P)_5TT$ also failed to induce triplex formation at 50 μM oligonucleotide (not shown).

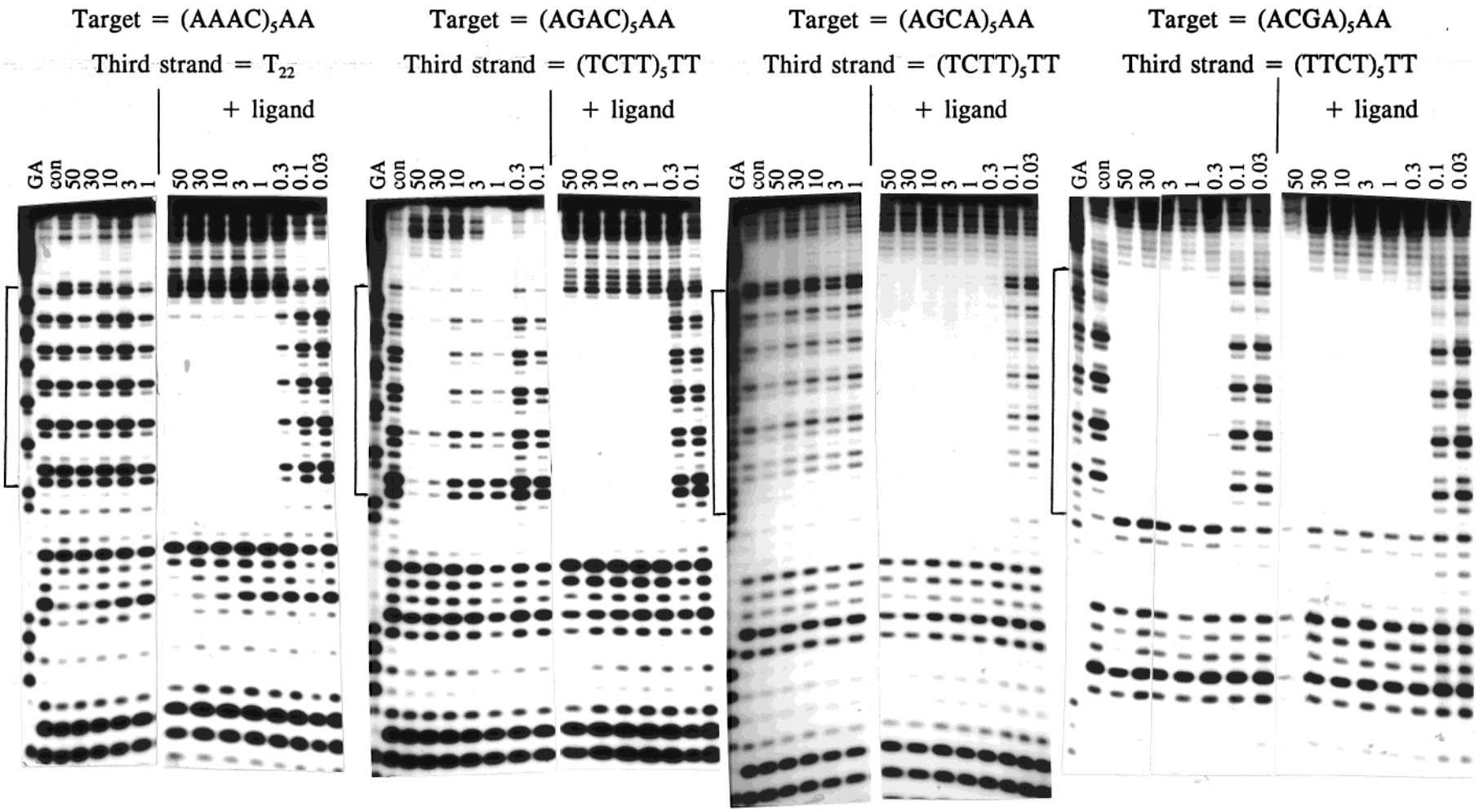

FIGURE 6: DNase I footprinting patterns showing the interaction of oligonucleotide T_{22} with fragment AAAC (left-hand panel), $(TCTT)_5TT$ with fragment AGAC (second panel), $(TCTT)_5TT$ with fragment AGCA (third panel), and $(TTCT)_5TT$ with fragment ACGA (right-hand panel). Oligonucleotide concentrations (μM) are shown at the top of each gel lane. In each panel, the right-hand lanes contained 10 μM naphthylquinoline triplex-binding ligand. Reactions were all performed in 50 mM sodium acetate, pH 5.5, containing 10 mM $MgCl_2$. Tracks labeled con are controls performed in the absence of added oligonucleotide or ligand. The tracks labeled GA are Maxam−Gilbert markers specific for purines. The square brackets show the positions of the target sites.

The final panel of Figure 6 shows the binding of $(TTCT)_5TT$ to the fragment ACGA. In this triplex, the T·CG triplet is flanked by T·AT on the 5′-side and C^+·GC on the 3′-side. In the absence of ligand, the oligonucleotide induces a clear footprint which persists to a concentration of about 0.3 μM. It appears that triplex formation at this site is more stable than at the other fragments containing CG inversions, suggesting that positioning a C^+·GC triplet on the 3′-side of T·CG improves its binding strength. In this instance, addition of the triplex-binding ligand has no effect on the footprint. Similarly, substitution of U^P for T using the oligonucleotide $(U^PTCU^P)_5TT$ also had no further effect on either the position or the concentration dependence of the footprint (not shown).

DISCUSSION

Increasing the range of sequences that can be targeted by triple helix formation, to include mixed-sequence DNA targets, is a major challenge in triplex recognition (*14, 15*). The work presented in this paper has addressed this problem by using unmodified and propargylamino-modified oligonucleotides to bind to target sequences which comprise nearly 25% pyrimidine residues.

G·TA-Containing Sites

Unmodified Oligonucleotides. Previous work has demonstrated that regions of alternating G·TA and T·AT triplets are not stable unless they are attached to an anchoring region of canonical triplets (*24, 26*). We have therefore investigated whether more stable structures could be induced by separating the G·TA triplets by more than one T·AT. We find that although three consecutive T·AT triplets are insufficient to induce triplex formation, placing five T·AT triplets between each G·TA enables triplex formation, albeit at an oligonucleotide concentration of 10 μM. Inspection of the right-hand panel of Figure 3 shows that although $(TTTTTG)_3TTTT$ generates a footprint with the target sequence $(AAAAAT)_3$-AAAA, cleavage products are still evident in the lower portion of the target site, suggesting that the oligonucleotide is not properly bound. This could reflect either fraying of the third strand at the end of the complex or formation of slipped structures in which the oligonucleotide only covers part of the target sequence. Addition of the naphthylquinoline triplex-binding ligand induces the formation of a footprint at AAAT and increases the apparent affinity at AAAAAT, reducing the oligonucleotide concentration required to generate a footprint and altering the footprint so that it now covers the entire target site. The effect of this ligand demonstrates that the inability to generate a footprint at AAAT is a result of the low affinity of the complex rather than any other inhibitory structural factors. The large increase in stability produced by this ligand contrasts with its inability to stabilize triplexes at $(AT)_n$ tracts (*24*), and is consistent with the suggestion that it binds between adjacent T·AT triplets. In addition, this arises from the low affinity of the complex and is not due to any other inherent characteristics.

Replacing the central T·AT triplet in each block of three T·ATs with C^+·GC (fragment AGAT) induces the formation of a complex which is stable in the absence of ligand. The large effect of introducing charged triplets shows the importance of positively charged species for generating stable triplexes and is consistent with previous studies which have shown that isolated C^+·GC triplets impart a greater stability than T·AT (*34*−*37*).

Comparison of triplex formation at fragments AGTA and ATGA, which have the same base composition as AGAT, reveals that the relative position of the C^+·GC and T·AT triplets affects triplex stability at these G·TA-containing sites. In this series of targets, the least stable complex is one in which the G·TA triplet is flanked by C^+·GC on the 5′-side. This is consistent with NMR studies which have suggested the formation of an additional inter-triplet hydrogen bond between G·TA and T·AT on its 5′-side (*19, 21*−*23*). Replacement of this T·AT with C^+·GC therefore results in a less stable complex. In contrast, replacement of the 3′-T·AT with C^+·GC does not seem to affect triplex formation and produces a complex with similar stability to that formed at AGAT. The destabilizing effect of a C^+·GC triplet on the 5′-side of G·TA therefore arises from removal of the extra interaction with a T·AT triplet rather than from the effect of the positive charge per se.

Modified Oligonucleotides. At the target site $(AAAT)_5AA$, incorporation of U^P residues in place of the central T in each block of three Ts leads to the formation of a triplex footprint with oligonucleotide concentrations as low as 0.3 μM in the absence of ligand. In this complex, the U^P·AT triplets occupy the same position as the C^+·GC triplets in the triplex at AGAT, and produce a structure with similar stability. The contribution of the positive charge on U^P presumably aids triplex formation by virtue of a favorable interaction with the phosphodiester backbone. Although it seems likely that the amino side chain of U^P interacts with the phosphate residues, magnesium is still essential for triplex formation with these modified oligonucleotides, since these 5 positive charges are not sufficient to screen the charge repulsion from the 22 third strand phosphates.

Experiments with fragment AGTA show that triplex stability is affected by the order of C^+·GC and U^P·AT triplets. This target, which produces the least stable of the C^+·GC-containing triplexes with unmodified oligonucleotides on account of positioning C^+·GC on the 5′-side of T·AT, shows a clear footprint with oligonucleotide $(TCGU^P)_5TT$. In this complex, the two charged triplets (C^+·GC and U^P·AT) alternate with uncharged triplets (T·AT and G·TA). In contrast, oligonucleotide $(U^PCGT)_5TT$, in which the C^+·GC and U^P·AT triplets are placed adjacent to each other, forms a less stable triplex than the unmodified oligonucleotide $(TCGT)_5TT$ at the target AGTA. It appears that placing the two charged triplets next to each other is therefore a very poor combination and significantly destabilizes the triplex. We presume that this effect arises from the proximity of positively charged groups, even though they are located in different positions on the bases. The triplex with $(U^PCGU^P)_5$-TT displays an intermediate stability, even though it contains three consecutive charged triplets. Although this was unexpected, it is consistent with the suggestion that adjacent U^P·AT triplets can generate very stable triplexes (*38*). It is important to note that adjacent C^+·GC triplets are destabilizing whereas adjacent U^P·AT triplets generate a favorable interaction. For C^+·GC, since the positive charge is located on the pyrimidine ring within the π-stacked bases, there will be electrostatic repulsion between adjacent charged residues which will also affect the p*K* values of contiguous cytosines.

In contrast, the positive charges on adjacent U^P•AT triplets will be neutralized by interaction with the phosphate groups. Adjacent U^P•AT triplets will also benefit from stacking of the propyne side groups. It is less clear why juxtaposition of U^P•AT and C^+•GC should be unfavorable. Comparison of the doubly U^P-substituted oligonucleotide at targets ATGA and AGTA reveals that the complex with the former requires slightly lower oligonucleotide concentrations than the latter, presumably because this complex does not possess the unfavorable C^+•GC on the 5′-side of G•TA

T•CG-Containing Sites

Unmodified Oligonucleotides. The equivalent series of DNA target sites containing T•CG triplets shows similar results to those containing G•TA triplets. However, in contrast to the G•TA-containing target AAAAAT, the sequence designed to form a triplex containing three T•CG triplets each separated by five T•ATs (fragment AAAAAC) does not generate a footprint unless the stabilizing ligand is present. This suggests that the T•CG triplet is less stable than G•TA. Similarly, introducing a C^+•GC triplet in fragment AGAC produces a triplex which only generates a DNase I footprint at the highest oligonucleotide concentrations, in contrast to the more stable G•TA-containing complex formed at AGAT. The presence of this charged triplet does not appear to be sufficient to promote the binding of this oligonucleotide. The interaction at each of these sites is potentiated by addition of the triplex-binding ligand, though this is somewhat surprising for sequence AGAC which forms a triplex which does not contain consecutive T•AT triplets. Since other studies have suggested that the naphthylquinoline triplex-binding ligand does not intercalate next to a C^+•GC triplet (*42*, *43*), this complex must be induced by interaction of the ligand between T•CG and T•AT. We have also been unable to stabilize blocks of alternating T•CG and T•AT triplets. In contrast to our results with alternating T•AT and G•TA triplets (*24*, *26*), we find that addition of a stabilizing tail does not promote the formation of alternating T•AT and T•CG, even in the presence of a triplex-binding ligand (unpublished observations).

As noted for the complexes containing G•TA, the position of the C^+•GC triplets has a pronounced effect on triplex stability. In this instance, the weakest complex is the one in which T•CG is flanked by C^+•GC on the 5′-side, as noted with G•TA. However, the strongest triplex contains a C^+•GC on the 3′-side of T•CG.

Modified Oligonucleotides. Replacing the central T•AT triplets with U^P•AT in the complex with fragment AAAC does not produce a large increase in stability, in contrast to the similar experiments with fragment AAAT. This again suggests that T•CG is less stable than G•TA. Similarly, neither $(U^PCTU^P)_5TT$ nor $(U^PTCU^P)_5TT$ shows any increase in binding to their respective target sites relative to the unmodified oligonucleotides. However, we cannot be sure whether this is due to a combination of effects from the two U^P residues which may have opposite effects on the stability of T•CG-containing triplexes

CONCLUSIONS

The results presented in this paper have shown that, by optimizing the arrangement of triplets, it is possible to achieve triplex formation at sites in which the targeted strand contains pyrimidine residues at every fourth base. However, triplex formation at sequences such as $(AT)_n$, $(AC)_n$, and $(ACAT)_n$ is still not possible, and achieving triplex formation at mixed-sequence DNA targets remains an elusive goal which will require the synthesis of novel base analogues.

REFERENCES

1. Soyfer, V. N., and Potaman, V. N. (1996) *Triple-Helical Nucleic Acids*, Springer-Verlag, Berlin, Heidelberg, and New York.
2. Thuong, N. T., and Hélène, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 666−690.
3. Moser, H. E., and Dervan, P. B. (1987) *Science 238*, 645−650.
4. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., and Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749−7760.
5. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 17−32.
6. Beal, P. A., and Dervan, P. B. (1991) *Science 251*, 1360−1363.
7. Radhakrishnan, I., and Patel, D. J. (1993) *Structure 1*, 135−152.
8. Radhakrishnan, I., de los Santos, C., and Patel, D. J. (1991) *J. Mol. Biol. 221*, 1403−1418.
9. Chen, F.-M. (1991) *Biochemistry 30*, 4472−4479.
10. Chan, P. P., and Glazer, P. M. (1997) *J. Mol. Med. 75*, 267−282.
11. Chubb, J. M., and Hogan, M. E. (1992) *Trends Biotechnol. 10*, 132−136.
12. Neidle, S. (1997) *Anti-Cancer Drug Des. 12*, 433−442.
13. Vasquez, K. M., and Wilson, J. H. (1998) *Trends Biochem. Sci. 23*, 4−9.
14. Doronina, S. O., and Behr, J. P. (1997) *Chem. Soc. Rev. 26*, 63−71
15. Gowers, D. M., and Fox, K. R. (1999) *Nucleic Acids Res. 27*, 1569−1577.
16. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967−971.
17. Kiessling, L. L., Griffin, L. C., and Dervan, P. B. (1992) *Biochemistry 31*, 2829−2834.
18. Fossella, J. A., Kim, Y. J., Shih, H., Richard, E. G., and Fresco, J. R. (1993) *Nucleic Acids Res. 21*, 4511−4515.
19. Chandler, S. P., and Fox, K. R. (1993) *FEBS Lett. 332*, 189−192.
20. Radhakrishnan I., Gao, X., de los Santos, C., Live, D., and Patel, D. J. (1991*) Biochemistry 30*, 9022−9030.
21. Radhakrishnan, I. Patel, D. J., and Gao, X. (1992) *Biochemistry 31*, 2514−2523.
22. Radhakrishnan, I., Patel, D. J. Veal, J. M., and Gao, X. (1992) *J. Am. Chem. Soc. 114*, 6913−6915.
23. Wang, E., Malek, S., and Feigon, J. (1992) *Biochemistry 31*, 4838−4846.
24. Chandler, S. P., and Fox, K. R. (1995) *FEBS Lett. 360*, 21−25.
25. Gowers, D. M., and Fox, K. R. (1997) *Nucleic Acids Res. 25*, 3787−3794.
26. Gowers, D. M., and Fox, K. R. (1998) *Nucleic Acids Res. 26*, 3626−3633.
27. Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., and Weis, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3840−3844.
28. Radhakrishnan, I., and Patel, D. J. (1994) *J. Mol. Biol. 241*, 600−619.
29. Kukreti, S., Sun, J. S., Loakes, D., Brown, D. M., Nguyen, C. H., Bisagni, E., Garestier, T., and Hélène,C. (1998) *Nucleic Acids Res. 26*, 2179−2183.
30. Zhou, B. W., Puga, E., Sun, J. S., Garestier, T., and Hélène, C. (1995) *J. Am. Chem Soc. 117*, 10425−10428.
31. Kukreti, S., Sun, J. S., Garestier, T., and Hélène, C. (1997) *Nucleic Acids Res. 25*, 4264−4270.

32. Miller, P. S., and Cushman, C. D. (1993) *Biochemistry 32*, 2999−3004.
33. Radhakrishnan, I., Patel, D. J., Priestley, E. S., Nash, H. M., and Dervan, P. B. (1993) *Biochemistry 32*, 11228−11234.
34. Roberts, R. W., and Crothers, D. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 4320−4325.
35. Keppler, M. D., and Fox, K. R. (1997) *Nucleic Acids Res. 25*, 4644−4649.
36. Volker, J., and Klump, H. H. (1994) *Biochemistry 33*, 13502−13508.
37. Asensio, J. L., Lane, A. N., Dhesi, J., Bergvist, S., and Brown, T. (1998) *J. Mol. Biol. 275*, 811−822.
38. Bijapur, J., Keppler, M. D., Bergvist, S., Brown, T., and Fox, K. R. (1999) *Nucleic Acids Res. 27*, 1802−1809.
39. Wilson, W. D., Tanious, F. A., Mizan, S., Yao, S. J., Kiselyov, A. S., Zon, G., and Strekowski, L. (1993) *Biochemistry 32*, 10614−10621.
40. Chandler, S. P., Strekowski, L., Wilson, W. D., and Fox, K. R. (1995) *Biochemistry 34*, 7234−7242.
41. Cassidy, S. A., Strekowski, L., Wilson, W. D., and Fox, K. R. (1994) *Biochemistry 33*, 15338−15347
42. Cassidy, S. A., Strekowski, L., and Fox, K.R. (1996) *Nucleic Acids Res. 24*, 4133−4138.
43. Moraru-Allen, A. A., Cassidy, S., Alvarez, J.-L. A., Fox, K. R., Brown, T., and Lane, A. N. (1997) *Nucleic Acids Res. 25*, 1890−1896.

BI9911637